REMARKS

Applicants and their Attorney thank the Examiner for the courtesy of several telephonic interviews (including November 6, November 8, November 28, November 29, 2006) during which the foregoing claim amendments and the outstanding rejections were discussed. The amendments/remarks presented herein are responsive to the Final Office Action dated June 1, 2006 and the Advisory Action dated January 5, 2007.

Applicants note that the present Preliminary Amendment incorporates arguments presented in the "Third Amendment and Response to Final Office Action," filed on December 1, 2007; the "Amendment and Response to Final Office Action," filed on October 2, 2006; the "Second Amendment and Response to Final Office Action," filed on November 9, 2006; and the "Supplemental Amendment," filed on November 28, 2006.

Claims 1-14, 70, and 71 were pending in the application. Claims 2-4, 6, 9-11, 13, and 70-71 have been cancelled, without prejudice. Claims 1, 5, 8, and 12 have been amended. Thus, upon entry of the present amendment, claims 1, 5, 7-8, 12, and 14 will remain pending in the application.

Support for the amendments to the claims may be found throughout the specification and the claims as originally filed. In particular, support for the amendments to claims 1 and 8 can be found, at least at, page 7, lines 27-29; page 10, lines 3-5; page 14, lines 19-24; page 68, lines 3-4; and page 70, lines 31-38 of the specification. In addition, support for the amendments to claims 1 and 8 can be found at least in Table 2B. In particular, the sequence corresponding to marker jlhbab412e01 (SEQ ID NO:16) is set forth on page 3 of Table 2B. Support for the amendments to claims 5 and 12 can be found, at least in, Tables 1-2B and at page 5, lines 3-5 and at page 10, lines 3-5 of the specification.

Any amendments to and/or cancellation of the claims was done solely for the purpose of expediting prosecution of the present application. Applicants reserve the right to pursue the subject matter of the claims as originally filed in this or a separate application(s).

Advisory Action Dated January 5, 2007

In the Advisory Action dated January 5, 2007, the Examiner indicated that the Amendment and Response filed on December 1, 2006 has not been entered because the proposed amendment to the claims to specify the marker jlhbab412e01 would allegedly raise new issues which would require further consideration and/or search. In particular, the Examiner is of the opinion that

> [a]ccording to Table 2A, 'jlhbab412e01' is an EST having the polynucleotide sequence of SEQ ID NO:16; but according to Table 5, it appears this same term 'jlhbab412e01' identifies more than one gene or gene product, which may or may not be distinct from the polynucleotide of SEQ ID NO:16, including: (a) Clone 741891 [RAB2, member RAS oncogene family-like//(Hs.170160; NM_004761)]; (b) Clone 122906 [ESTs//(Hs. 186545)]; (c) Clone 1946534 [lymphotoxin beta (TNF superfamily, member 3)//(Hs.890; NM_002341)]; (d) Clone 837891 [UNIGENE-ambiguity: Hs.271869::Hs.267654! ESTs//(Hs.271869)]; and (e) Clone 1493205 [ESTs, Weakly similar to WD40 protein Ciao 1 [H. sapiens]//(Hs.90680)]. Therefore, where the elected invention is the invention of Group I, a method for determining whether an agent can or cannot reduce the growth of at tumor comprising determining whether the tumor cells express the sensitivity marker of SEQ ID NO:16, it is not apparent to what extent the claims read on this elected invention.

Contrary to the Examiners' assertions, Applicants submit that Table 5 does <u>not</u> disclose that jlhbab412e01 identifies more than one gene or gene product. Specifically, as indicated by a copy of Table 5 attached herein as Appendix B, the right column of the row containing marker jlhbab412e01 in Table 5 has been left blank, indicating that a description of jlhbab412e01 was not available at the time of filing the instant application. Hence, jlhbab412e01 was a novel marker and had not previously been described in the art. The Examiner refers to several clones, or assigned reference numbers for specific sequences, listed below jlhbab412e01, including 741891, 122906, 1946534, 837891, and 1493205. Applicants wish to point out that ***these clones are distinct from jlhbab412e01***. Each of these clones had previously been identified, and the brief description of each clone follows in the right column of the same row of the clone. For example, "1946534" refers to a sequence that had been previously described as lymphotoxin beta (TNF superfamily, member 3), which is encoded by UniGene reference number Hs.890 and Genbank Accession number NM_002341. Thus, jlhbab412e01 is a novel marker identified in

the instant application and does not correspond to the clones 741891, 122906, 1946534, 837891, and 1493205 which are listed below jlhbab412e01.

The Examiner is also of the opinion that "it is not apparent whether this marker is necessarily a mRNA molecule comprising the corresponding RNA sequence of the polynucleotide of SEQ ID NO: 16." Applicants submit that the term "marker" is defined in the specification at least at page 3, line 36 through page 4, line 7 as

> **a naturally-occurring polymer corresponding to at least one of the nucleic acids listed in Tables 1-6**. For example, markers include, without limitation, sense and anti-sense strands of genomic DNA (*i.e.* including any introns occurring therein), RNA generated by transcription of genomic DNA (*i.e.* prior to splicing), RNA generated by splicing of RNA transcribed from genomic DNA, and proteins generated by translation of spliced RNA (*i.e.* including proteins both before and after cleavage of normally cleaved regions such as transmembrane signal sequences). As used herein, "marker" may also include a cDNA made by reverse transcription of an RNA generated by transcription of genomic DNA (including spliced RNA). (Emphasis added).

Based on the foregoing teachings in Applicants' specification and the general knowledge in the art, the ordinarily skilled artisan would understand that the term "marker" refers to any naturally-occurring polymer corresponding to the nucleic acid sequence of SEQ ID NO:16, as defined above. With respect, in particular, to dependent claims 5 and 12, Applicants submit that it is apparent that these claims are specifically directed to one such naturally-occurring polymer, *e.g.*, "an mRNA molecule comprising the corresponding RNA sequence of the polynucleotide sequence of SEQ ID NO:16".

Advisory Action Dated October 18, 2006

In the Advisory Action dated October 18, 2006, the Examiner has indicated that the Amendment and Response filed on October 2, 2006 has not been entered because the amendments to claims 5 and 12 would raise new issues under 35 U.S.C. §112, first and/or second paragraph. According to the Examiner, "SEQ ID NO:16, however, is the polynucleotide sequence of a complementary DNA (cDNA) molecule; and no such mRNA transcript would exist in the sample."

Responsive to the Examiner's comments, Applicants present herein new amendments to claims 5 and 12, thereby rendering the foregoing issues raised by the Examiner moot. Specifically, claims 5 and 12 have been amended to specify that the amount of "an mRNA molecule comprising the corresponding RNA sequence of the polynucleotide sequence of SEQ ID NO:16" is detected. Applicants respectfully note that the claims have been amended based upon the Examiner's suggestions (see page 12 of the Final Office Action), however, Applicants have slightly modified the suggested claim language in order to be more scientifically accurate.

As the Examiner knows, "[t]ranscription involves synthesis of an RNA chain representing one strand of a DNA duplex," *i.e.*, the RNA transcript is *identical in sequence (except for the substitution of U for T)* to the exonic sequences of the DNA coding strand. This concept is graphically depicted in Figure 14.1 of Genes V, in which the green RNA transcript is identical to the green DNA coding strand (see Figure 14.1 of Genes V, Edition of 1994, see page 377 in Appendix A submitted herewith).

Reverse transcription makes it possible to synthesize a duplex DNA from any mRNA molecule (see Figure 21.5, of Genes V, see page 641 in Appendix A). First, a primer is annealed to the poly(dA) tail of the mRNA. The enzyme, reverse transcriptase, engages in the usual 5'-3' elongation, adding deoxynucleotide one at a time, as directed by complementary base pairing with the mRNA template (described in detail at page 641). A reaction occurs at the end of the mRNA, in which the enzyme causes the reverse transcript to 'loop back' on itself, by using the last few bases of the reverse transcript as a template for synthesis of a complement. The product of the reaction is a hybrid molecule, consisting of a template RNA strand base-paired with the complementary DNA strand. The original mRNA is then degraded by treatment with alkali and **the product is a single stranded DNA that is complementary to the mRNA; it is called cDNA** (see page 642 of Genes V). The hairpin at the 3' end of the cDNA provides a natural primer for the next step, the use of *E. coli* DNA polymerase I to convert the single-stranded cDNA into a duplex DNA. In this reaction, the enzyme uses the cDNA as template **for synthesis of a sequence identical with the original mRNA except for the substitution of T in the DNA for U in the original RNA** (page 642). This is called a **cDNA clone**. (From the terminology, a somewhat looser use of the term 'cDNA' has emerged, being taken to describe the duplex insert and not just the original single-stranded reverse transcript) (page 642).

For the Examiner's convenience, Applicants provide below a simplified representation of the processes described above.

5'-ATGCGCCAT-3' [DNA coding strand]

3'-TACGCGGTA-5' [DNA Non-coding strand/template strand]] DNA

↓ Transcription

5'-AUGCGCCAU-3' [mRNA sequence is complementary to DNA non-coding strand and identical to the DNA coding strand]] mRNA

↓ Reverse Transcription

5'-ATGCGCCAT -3' [Coding strand, *e.g.*, SEQ ID NO:16]

3'-TACGCGGTA-5' [Non-coding strand]] cDNA

As evidenced by all of the foregoing, the cDNA coding strand, *e.g.*, SEQ ID NO:16, is identical to the mRNA sequence (except for the substitution of U with T); it is not the complement of the mRNA sequence. Thus, Applicants respectfully request that the amended claim language "an mRNA molecule comprising the corresponding RNA sequence of the polynucleotide sequence of SEQ ID NO:16" be considered and entered by the Examiner.

The foregoing amendments were made solely in the interest of expediting prosecution and allowance of the present application and do not in any way reflect an acquiescence to any of the Examiner's rejections.

Objections to the Oath

The Examiner has objected to the oath/declaration as being defective for having non-initialed and/or non-dated alterations.

Applicants are in the process of contacting Edwin Clark to obtain a new oath/declaration in compliance with 37 C.F.R. §1.67(a). Applicants will submit the new oath/declaration as soon as it is available and certainly prior to the issuance of the present application.

Elections/Restrictions

The Examiner is of the opinion that "claims 1, 5-8, 11-14, 70 and 71 are directed to an invention(s) that is (are) independent or distinct from the invention originally claimed" (page 3 of the Office Action). According to the Examiner, the elected invention was drawn to

> a process for determining whether an agent can be used to reduce the growth of a tumor, said process comprising obtaining a sample of tumor cells and determining whether the tumor cells express one or more sensitivity markers, wherein said 'agent' is a combination of a taxane compound and a platinum compound.

The Examiner objects to the claims on the grounds that they are directed to any agent selected from the group consisting of TAXOL, TAXOL mimics, TAXOL analogs, TAXOL derivatives, cisplatin, cisplatin mimics, cisplatin analogs, and cisplatin derivatives.

Applicants acknowledge the election of the species of SEQ ID NO:16 and, further, the species election of a combination of agents consisting of a taxane compound and a platinum compound. Responsive to the Examiner's comments, Applicants have cancelled claims 2-3 and 9-10, without prejudice, as being directed to a non-elected invention. In the interest of expediting prosecution, and in no way acquiescing to the Examiner's objection, Applicants have also cancelled claims 4 and 11, without prejudice. Moreover, Applicants have amended claims 1 and 8 to specify that the claimed agent is "paclitaxel and cisplatin."

It is Applicants' understanding that upon the allowance of a generic claim, Applicants will be entitled to consideration of claims to additional species that are written in dependent form or otherwise include all the limitations of an allowed generic claim as provided by 37 C.F.R. §1.141 *et seq.* Claims 1, 5-8, 12-14, and 70-71 read on the elected species of a combination of agents consisting of a taxane compound and a platinum compound. Claims 1, 5, 6, 8, 12, 13, 70, and 71 read on the elected species of SEQ ID NO:16.

Objection to the Specification

The Examiner has objected to the specification for containing embedded hyperlinks and/or other form of browser-executable code and for the use of improperly demarcated trademarks.

Responsive to the Examiner's objection, Applicants have submitted a Substitute Specification in which such hyperlinks have been removed and the improperly demarcated

trademarks have been corrected. Accordingly, Applicants respectfully request reconsideration and withdrawal of this objection.

Rejection of Claims 1, 4-8, 11-14, 70, and 71 Under 35 USC § 112, First Paragraph

The Examiner has rejected claims 1, 4-8, 11-14, 70, and 71 under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement. The Examiner's grounds for rejection are two-fold and Applicants will address each issue in turn.

Firstly, the Examiner asserts that "the claims are directed to a variety of structurally and/or functionally distinct compounds, which are not necessarily 'taxanes' or 'platinum compounds,' per se" (see page 10 of the Office Action). The Examiner concludes that "...the specification would not permit the skilled artisan [to] immediately envision, recognize or distinguish at least a substantial number of these compounds, so as to recognize that Applicant did in fact have possession of the claimed invention at the time the application was filed."

Applicants respectfully traverse the aforementioned rejection and respectfully submit that there is sufficient written description in Applicants' specification regarding the claimed methods, and that "the description clearly allow[s] persons of ordinary skill in the art to recognize that [Applicants] invented what is claimed" (see M.P.E.P. §2163.02). In particular, the present specification teaches that TAXOL and cisplatin are chemical compounds within the family of taxane and platinum compounds, respectively, which are art-recognized as families of *related* compounds (page 66, lines 4-5; and page 69, line 34 through page 70, line 1 of the specification) and Applicants further describe identifying characteristics of the members of these families, including the structure of analogs and derivatives that are structurally or functionally similar to TAXOL or cisplatin (see page 66, line 3 through page 70, line 24 of the specification).

However, in the interest of expediting prosecution, and in no way acquiescing to the validity of the Examiner's rejection, Applicants have amended claims 1 and 8 such that they are now directed to "***paclitaxel and cisplatin***," thereby rendering the foregoing rejection moot.

Secondly, the Examiner is of the opinion that the specification does not sufficiently describe the chemical structure of SEQ ID NO:16 (see pages 11-14 of the Office Action), namely, the specification does not sufficiently describe the genomic DNA molecule corresponding to SEQ ID NO:16 (11-12 of the Office Action) or the structure of the protein encoded by SEQ ID NO:16 (see page 12 and 14 of the Office Action).

Applicants respectfully traverse the rejection for the reasons of record. In addition, Applicants wish to make the following remarks of record. A description of the corresponding genomic sequence is not required to practice the claimed invention. Based on the nucleotide sequence of a marker of the invention, such as, for example, the marker of SEQ ID NO:16 (depicted in Tables 2A-2B as "jlhbab412e01"), one of skill in the art would readily be able to design probes or primers suitable for detecting the expression of this marker (as described at, for example, page 15, lines 30-35 and page 17, lines 12-19 of the specification). Moreover, it is well known in the art that the amino acid sequence corresponding to a nucleic acid sequence may be determined by identifying the open reading frame or translating the sequence in all three reading frames using well known programs designed to identify open reading frames and/or translate the nucleotide sequence in all frames. Therefore, based on the disclosure of the nucleotide sequences of the markers of the invention and the teaching in the specification that these sequences represent expressed products, one of skill in the art would conclude that Applicants were in possession of the corresponding protein sequence at the time of filing.

However, Applicants respectfully submit that in the interest of expediting prosecution, and in no way acquiescing to the validity of the Examiner's rejection, Applicants have cancelled claims 6 and 13, thereby rendering the foregoing rejection moot.

In view of the foregoing, Applicants respectfully request that the aforementioned rejection be reconsidered and withdrawn.

Rejection of Claims 1, 4-8, 11-14, 70, and 71 Under 35 USC § 112, First Paragraph

The Examiner has rejected claims 1, 4-8, 11-14, 70, and 71 under 35 U.S.C. §112, first paragraph, as failing to comply with the enablement requirement. ***The Examiner admits that "the specification teaches a correlation between the presence of one or more adequately described markers in ovarian cancer cells and their sensitivity or lack thereof to a combination of Taxol™ (paclitaxel) and cisplatin"*** (emphasis added). However, the Examiner is of the opinion that "[t]he combination of Taxol™ and cisplatin is not representative of the whole of the genus of agents to which the claims are directed" (page 16 of the Office Action).

Applicants traverse the foregoing rejection for the reasons of record. It is Applicants' position that the amount of direction and guidance disclosed in the specification is

commensurate with the scope of the claims and sufficient to enable the skilled artisan to make and use the claimed methods using only routine experimentation.

However, in the interest of expediting prosecution, and in no way acquiescing to the validity of the Examiner's rejection, Applicants have amended claims 1 and 8 such that they are now directed to "***paclitaxel and cisplatin***." The Examiner has admitted in the present Office Action that the specification is enabling for such methods (see above). Thus, Applicants respectfully submit that the foregoing rejection has been rendered moot and request that the Examiner reconsider and withdraw this rejection.

The Examiner has also questioned why "...the absence or underexpression of the one or more markers and the tumor cells' insensitivity to an agent are inversely correlated, whereas it is merely the presence of the marker, rather than its overexpression that allegedly positively correlates with the tumor cells' sensitivity to the agent." The Examiner further asserts that "[t]he specification fails to teach whether it is the mere presence of such markers, or their relative levels of expression that correlate with tumor cells' sensitivities to agents" (page 20 of the Office Action).

Applicants respectfully traverse this rejection. With respect to the ***absence versus the presence of a marker***, the Applicants teach, at page 70, lines 31-38 of the specification, that

> [i]f the gene is expressed, and the marker of the invention to which the gene corresponds is a sensitivity marker, then the therapeutic agent will be effective against the cancer. Accordingly, if a sensitivity marker is not expressed, then the therapeutic agent will not be effective against the cancer. If a resistance marker of the invention is expressed, then the therapeutic agent will not be effective against the cancer. Accordingly, if the resistance marker is not expressed, then the therapeutic agent will be effective against the cancer.

With respect to the ***relative level of expression of a marker***, Applicants teach, at page 71, lines 3-13 of the specification, that

> [b]y examining the expression of one or more of the identified markers in a sample of cancer cells taken from a patient during the course of therapeutic treatment, it is also possible to determine whether the therapeutic agent is continuing to work or whether the cancer has become resistant (refractory) to the treatment protocol. For example, a cancer patient receiving a treatment of TAXOL would have cancer cells removed and monitored for the expression of a marker. If the expression level of a sensitivity marker remains substantially the same, the treatment with TAXOL would continue. However, a significant decrease in sensitivity

> marker expression or increased expression of a resistance marker, would suggest that the cancer may have become resistant to TAXOL and another chemotherapy protocol should be initiated to treat the patient.

Furthermore, the specification provides a ***description of the standard to which such comparisons of the levels of expression are to be made***. For example, at page 6 lines 13-21 of the specification, expression levels are defined a follows:

> [e]xpression of a marker in a patient is 'significantly' higher or lower than the normal level of expression of a marker if the level of expression of the marker is greater or less, respectively, than the normal level by an amount greater than the standard error of the assay employed to assess expression, and preferably at least twice, and more preferably three, four, five or ten times that amount. Alternately, expression of the marker in the patient can be considered 'significantly' higher or lower than the normal level of expression if the level of expression is at least about two, and preferably at least about three, four, or five times, higher or lower, respectively, than the normal level of expression of the marker.

Applicants also teach the use of normalization by comparing the expression of a marker of the instant invention "to the expression of a gene that is not a sensitivity or resistance gene, *e.g.*, a housekeeping genes that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene" (page 14, lines 34-37 of the specification).

In an effort to expedite prosecution and in no way conceding the validity of the Examiner's rejection, Applicants have amended claim 8 to specify "one or more of the sensitivity markers in Tables 1-6 is not expressed by the ovarian tumor cells," thereby rendering the foregoing rejection moot.

Based on the foregoing, Applicants respectfully request reconsideration and withdrawal of this rejection under 35 U.S.C. §112, first paragraph.

New Grounds of Rejection

Rejection of Claims 1, 4-8, 11-14, 70, and 71 Under 35 USC § 112, Second Paragraph

The Examiner has rejected claims 1, 4-8, 11-14, 70, and 71 under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as the invention. In particular, the Examiner is of the

opinion that "[t]he claim scope is uncertain since the trademark or trade name cannot be used properly to identify any particular material or product."

In an effort to expedite prosecution and in no way conceding the validity of the Examiner's rejection, Applicants have amended claims 1 and 8 to recite "paclitaxel," thereby rendering the foregoing rejection moot.

The Examiner is further of the opinion that claims 5 and 12 are indefinite because "[t]he polynucleotide sequence of SEQ ID NO:16 is not the sequence of an RNA molecule; and the term 'SEQ ID NO:16 mRNA' is not defined in the specification."

Without acquiescing to this rejection and solely in an effort to further prosecution, Applicants have amended claims 5 and 12 to specify that the amount of "an mRNA molecule comprising the complement of the polynucleotide sequence of SEQ ID NO:16" is detected, as suggested by the Examiner (see page 12 of the Final Office Action). In view of the foregoing, Applicants respectfully request reconsideration and withdrawal of this rejection under 35 U.S.C. §112, second paragraph.

CONCLUSION

In view of the foregoing, entry of the amendments and remarks presented, favorable reconsideration and withdrawal of the rejections, and allowance of this application with the pending claims are respectfully requested. If a telephone conversation with the Applicants' attorney would expedite prosecution of the above-identified application, the Examiner is invited to call the undersigned at (617) 227-7400.

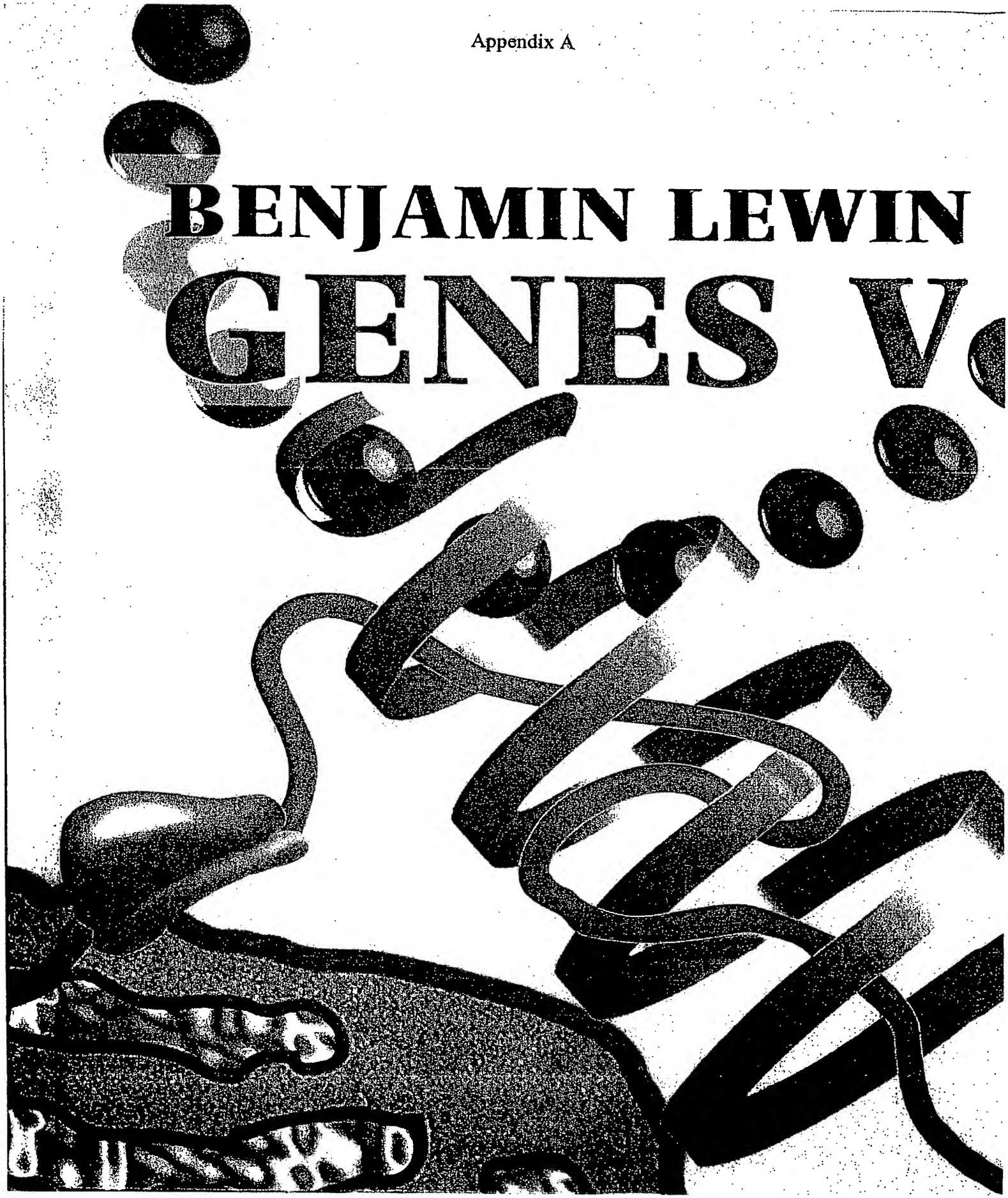
Dated: June 28, 2007

Respectfully submitted,

By _____

Maria Laccotripe Zacharakis, Ph.D., J.D.
Registration No.: 56,266
LAHIVE & COCKFIELD, LLP
One Post Office Square
Boston, Massachusetts 02109
(617) 227-7400
(617) 742-4214 (Fax)
Attorney/Agent For Applicants

Appendix A



GENES V

Benjamin Lewin

OXFORD UNIVERSITY PRESS
Oxford New York Tokyo
1994

CH

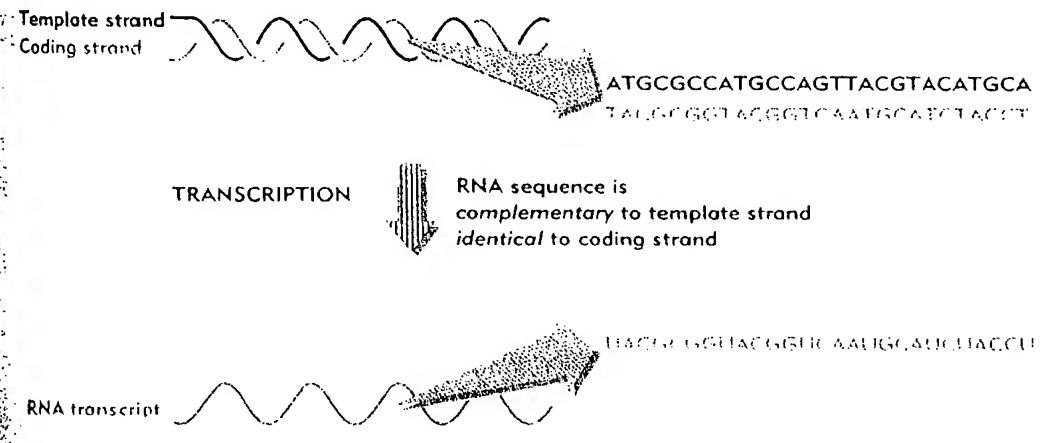
Control at initiation: RNA polymerase–promoter interactions

Transcription involves synthesis of an RNA chain representing one strand of a DNA duplex. By 'representing' we mean that the RNA is *identical in sequence* with one strand of the DNA, which is called the **coding strand**. It is *complementary* to the other strand, which provides the template for its synthesis. This relationship between double-stranded DNA and its single-stranded RNA transcript is recapitulated in Figure 14.1.

RNA synthesis is catalyzed by an enzyme, called **RNA polymerase**. Transcription starts when RNA polymerase binds to a special region, the **promoter**, at the start of the gene. The promoter surrounds the first base pair that is actually transcribed into RNA, the **startpoint**. From this point, RNA polymerase moves along the template, synthesizing RNA, until it reaches a **terminator** sequence. This action defines a **transcription unit** that extends from the

Figure 14.1

Overview: the function of RNA polymerase is to copy one strand of duplex DNA into RNA.



any additional material between them. A problem inherent in this technique is that there is no control over which pairs of blunt ends are joined together, so it is necessary first to perform the reaction and then to isolate the desired products from among the other products.

There are numerous variations of these methods. One technique uses short DNA duplexes ('linkers') that contain the *Eco*RI (or some equivalent) recognition palindrome. The linkers can be synthesized chemically, and are added covalently to the ends of a plasmid or an insert by blunt-end ligation. The inserted DNA can be retrieved by cleavage with *Eco*RI, but there are no restrictions on the original choice of sites to generate the ends. With sufficient manipulation, it is now possible to insert any foreign DNA fragment into any particular vector site, and to arrange for retrieval of the fragment when necessary.

When a foreign DNA fragment is inserted into a plasmid, it can be connected in either orientation, that is, with either of the ends of the foreign DNA joined to either of the ends of the plasmid. This does not matter when the purpose of cloning is simply to amplify the inserted sequence. However, it is important when the experiment is designed to obtain expression of the foreign DNA, which requires insertion in a particular orientation.

In this case, populations of plasmids carrying the plasmid in either orientation are obtained by random insertion, after which they are characterized by restriction mapping to identify the desired class. Or the experiment is designed so as to permit insertion in one orientation only. For example, each of the DNAs, vector and insert, can be cleaved with *two* restriction enzymes that yield different sticky ends, to generate the type of pattern where each DNA has the sequence

```
End 1 ——————————End 2
```

Now if only the two end-1 sequences can anneal together, and only the two end-2 sequences can anneal, the insertion can take place only in one orientation, generating the chimeric plasmid

```
End 1———Insert———End 2
  |                  |
End 1———Plasmid———End 2
```

Copying mRNA into cDNA

One of the principal uses of cloning technology is to isolate specific genes directly from the genome. Any particular gene represents only a very small part of a eukaryotic genome. In a typical mammal, the size of the genome is ~10^9 bp, so that a single gene of (say) 5000 bp represents only 0.0005% of the total nuclear DNA.

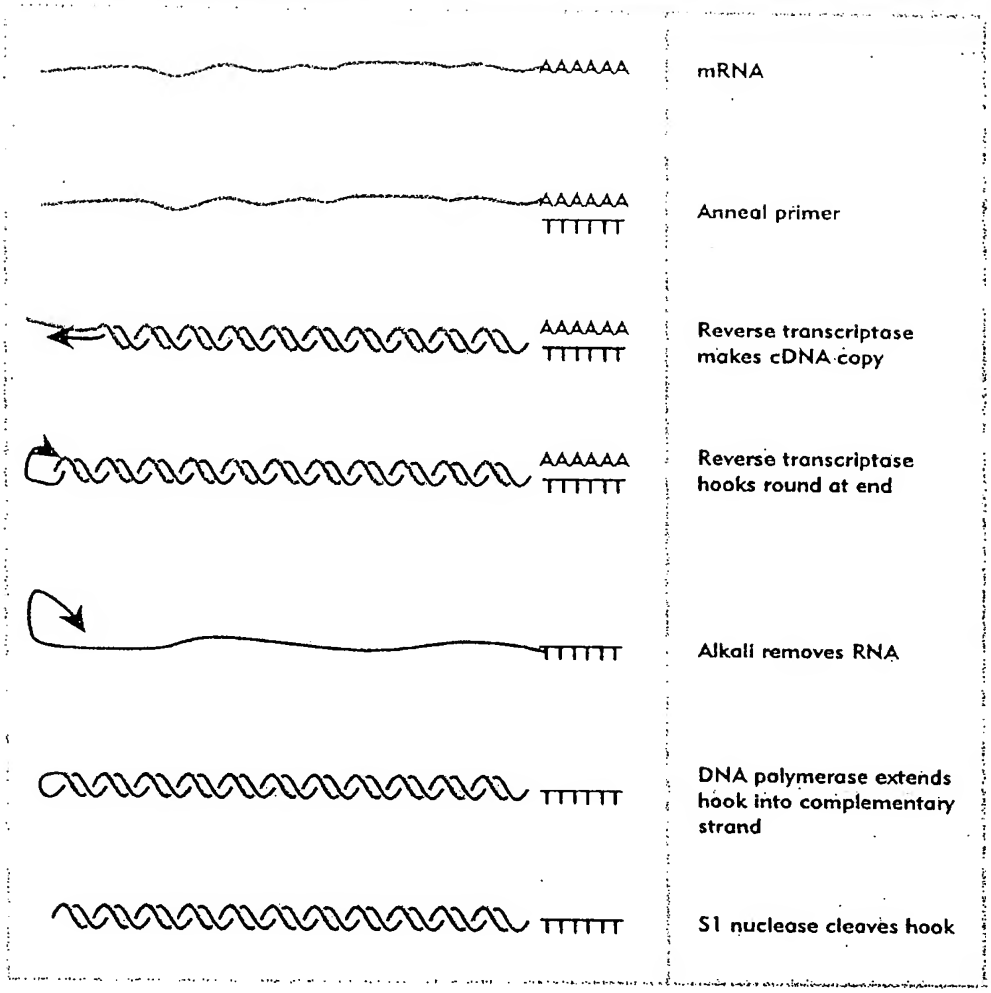
To identify such a tiny proportion, we need a very specific probe that reacts *only* with the particular sequence in which we are interested, to pick it out from the vast excess of other sequences. The usual technique is to use a highly labeled radioactive probe of RNA or DNA, whose hybridization with the gene is assayed by autoradiography.

For the purpose of obtaining a DNA sequence that represents a particular protein, the place to start is with mRNA, which, after all, is the template used to produce the protein *in vivo*. But it can be difficult to obtain the mRNA that represents a particular protein when the product is rare. There are several techniques for isolating an mRNA via the properties of its product, but a common problem is the requirement that the RNA must first be purified.

Rather than purify the RNA, a DNA copy of the RNA sequence is made. This has the great advantages that unlimited amounts of material can be obtained, and the DNA can be radioactively

Figure 21.5

mRNA can be copied into double-stranded DNA.



eled, providing a much more powerful probe. The existence of reverse transcription makes it possible to synthesize a duplex DNA from any mRNA. This is especially easy for mRNAs that carry a poly(A) tail at the 3′ end, as illustrated in Figure 21.5. First, a primer is annealed to the poly(dA). It is a short sequence of oligo(dT), whose purpose is to provide a free 3′ end that can be used for extension by the enzyme reverse transcriptase. The enzyme engages in the usual 5′–3′ elongation, adding deoxynucleotides one at a time, as directed by complementary base pairing with the mRNA template. The product of the reaction is a hybrid molecule, consisting of a template RNA strand base-paired with the complementary DNA strand. The only practical problem is the propensity *in vitro* of reverse transcriptase to stop before it has reached the 5′ end of the mRNA. In this case, the resulting reverse transcript falls short of representing the entire mRNA, because it lacks some of the sequences complementary to the 5′ end. However, by judicious adjustment of the experimental conditions, usually it is possible to persuade reverse transcriptase to proceed all the way.

A useful reaction tends to occur at the end of the mRNA, where the enzyme causes the reverse

transcript to 'loop back' on itself, by using the last few bases of the reverse transcript as a template for synthesis of a complement. That is, the end of the complementary DNA is used to direct synthesis of a short sequence that is identical with the mRNA, and which displaces it. This creates a short hairpin, usually 10–20 bp long.

At this juncture, the original mRNA is degraded by treatment with alkali (a procedure that does not affect DNA). The product is a single-stranded DNA that is complementary to the mRNA; it is called cDNA.

The hairpin at the 3′ end of the cDNA provides a natural primer for the next step, the use of *E. coli* DNA polymerase I to convert the single-stranded cDNA into a duplex DNA via synthesis of the complementary strand. In this reaction, the enzyme uses the cDNA as template for synthesis of a sequence identical with the original mRNA. The product is a duplex molecule with a hairpin at one end. The hairpin is cut by the enzyme S1 nuclease (which specifically degrades single-stranded DNA) to generate a conventional DNA duplex.

The duplex DNA can be cloned to generate large amounts of a synthetic gene representing the mRNA sequence. This is called a **cDNA clone**. (From this terminology, a somewhat looser use of the term 'cDNA' has emerged, being taken to describe the duplex insert and not just the original single-stranded reverse transcript.)

The power of sequencing technology has made it possible to bypass most of the problems posed by rare mRNAs by targeting a probe directly to the mRNA sequence. One powerful technique requires knowledge of only a small sequence of the protein. Short oligonucleotides can be synthesized that correspond to this sequence. A variety of oligonucleotides can be made to cover possible alternative codons, especially at third base positions. These oligonucleotides can be used to isolate cDNAs or genomic DNA that include the sequence of the corresponding gene.

Isolating individual genes from the genome

The first step toward identifying the gene corresponding to a particular probe is to break the DNA of the genome into fragments of a manageable size. It is desirable to obtain the gene in as few fragments as possible (ideally only one). Usually the maximum lengths of DNA that can be manipulated directly are in the range of 15–20 kb. Sometimes it is not possible to obtain a gene in the form of a single fragment, and then its structure must be determined by piecing together the information gained from its various fragments. (We have discussed the use of overlapping fragments in Chapter 6.)

The best technique for fragmenting a genome is to make a restriction digest. Then every fragment ends in a site that was recognized by that particular enzyme. However, restriction sites may occur at inconvenient locations—for example, in the middle of a gene that is to be cloned. One way to avoid this is to use more than one restriction enzyme, or to repeat the experiment with different enzymes whose recognition sites lie at different locations. But this is time consuming; and when a long sequence is involved, it may be difficult to find an enzyme that does not cleave within it.

When the DNA of an entire genome is digested with a restriction enzyme, the frequency of breakage is controlled by the length of the sequence recognized by the enzyme. The longer the sequence, the less often it occurs by chance. The probability that a particular 4 bp sequence will occur is $0.25^4 = 1/256$, so that an enzyme with such a short recognition sequence will cleave DNA rather frequently. The frequency declines to 1/1000 for a 5 bp sequence and to 1/4000 for a 6 bp sequence.

(This calculation assumes that each base

Appendix B
Table 5

Clone	Annotation
242642	EST//(Hs.42041;)
121661	ESTs//(Hs.226410;)
739193	cellular retinoic acid-binding protein 1//(Hs.7678;NM_004378)
32567	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein//(Hs.34853;NM_001546)
788234	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein//(Hs.34853;NM_001546)
809694	cellular retinoic acid-binding protein 1//(Hs.7678;NM_004378)
27544	prominin (mouse)-like 1//(Hs.112360;NM_006017)
809998	UNIGENE-ambiguity: Hs.252475::Hs.250817! amylase, alpha 2B; pancreatic//(Hs.252475;)
789369	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein//(Hs.34853;NM_001546)
1412238	amylase, alpha 2A; pancreatic//(Hs.75733;NM_000699)
755881	ESTs//(Hs.179902;)
743465	ESTs//(Hs.112703;)
460487	lactotransferrin//(Hs.347;NM_002343)
1609538	ESTs//(Hs.104696;)
298417	trefoil factor 3 (intestinal)//(Hs.82961;NM_003226)
1412245	carboxypeptidase A2 (pancreatic)//(Hs.89717;NM_001869)
322723	ESTs//(Hs.93231;)
jlhbab397f01	
jlhbae334b03	
154654	ESTs, Highly similar to IROQUOIS-CLASS HOMEODOMAIN PROTEIN IRX-3 [M.musculus]//(Hs.3321;)
842863	N-myc downstream regulated//(Hs.75789;NM_006096)
1473682	DKFZP586G1624 protein//(Hs.125262;)
285507	EST//(Hs.161495;)
277173	laminin receptor 1 (67kD, ribosomal protein SA)//(Hs.181357;NM_002295)
jlhbab453e07	
488945	UNIGENE-ambiguity: Hs.252475::Hs.180149! amylase, alpha 2B; pancreatic//(Hs.252475;)
jlhbab412e01	
741891	RAB2, member RAS oncogene family-like//(Hs.170160;NM_004761)
122906	ESTs//(Hs.186545;)
1946534	lymphotoxin beta (TNF superfamily, member 3)//(Hs.890;NM_002341)
837891	UNIGENE-ambiguity: Hs.271869::Hs.267654! ESTs//(Hs.271869;)
1493205	ESTs, Weakly similar to WD40 protein Ciao 1 [H.sapiens]//(Hs.90680;)
jrhoc127f11	
jlhbac238e10	
713263	prepronociceptin//(Hs.89040;)
970649	ESTs//(Hs.116561;)
jlhbad283g07	
767993	Homo sapiens mRNA; cDNA DKFZp586L2123 (from clone DKFZp586L2123)//(Hs.29759;)

Appendix B
Table 5

1456937	oviductal glycoprotein 1, 120kD//(Hs.1154;NM_002557)
284220	CD22 antigen//(Hs.171763;NM_001771)
1635978	ESTs//(Hs.131201;)
731311	ESTs, Weakly similar to ORF YKL201c [S.cerevisiae]//(Hs.270266;)
jMhXp229h07	
785847	ubiquitin-conjugating enzyme E2M (homologous to yeast UBC12)//(Hs.200478;NM_003969)
788524	ESTs, Weakly similar to KIAA1006 protein [H.sapiens]//(Hs.99291;)
759163	microfibrillar-associated protein 4//(Hs.118223;)
788609	ESTs, Weakly similar to similar to KIAA0766 [H.sapiens]//(Hs.213586;)
128126	decay accelerating factor for complement (CD55, Cromer blood group system)//(Hs.1369;)
1635203	ESTs, Weakly similar to weak similarity to collagens [C.elegans]//(Hs.127824;)
859858	steroidogenic acute regulatory protein//(Hs.3132;NM_000349)
jrhob001h03	
431944	ESTs//(Hs.117106;)
815284	peptidase D//(Hs.73947;NM_000285)